

CLAIMS

1. (Currently Amended) A packet voice conferencing method comprising:
 - receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;
 - digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;
 - estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
 - transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
 - wherein estimating the relative temporal delay further comprises calculating, for each of a plurality of relative time shifts, a first-to-second sound field signal cross-correlation coefficient, selecting the relative temporal delay to correspond to the relative time shift generating the largest cross-correlation coefficient, and tracking the beginning and ending of a talkspurt represented in the sound field signals, and limiting the variation of the estimated relative temporal delay during a talkspurt.
2. (Original) The method of claim 1, wherein digitally encoding a signal block comprises combining the first and second sound field signals into a composite sound field signal by a method selected from the group of methods consisting of:
 - selecting one sound field signal as the source of the composite sound field signal and discarding the other sound field signal;
 - summing the first and second sound field signals; and
 - averaging the first and second sound field signals.
3. (Canceled)
4. (Canceled)
5. (Original) The method of claim 1, wherein the relative temporal delay associated with the first time period is estimated using substantially only the sound field signals captured during the first time period.

6. (Previously Presented) A packet voice conferencing method comprising:

- receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;
- digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;
- estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
- transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
- wherein estimating the relative temporal delay further comprises tracking the beginning and ending of a talkspurt represented in the sound field signals, wherein relative temporal delay associated with the first time period is estimated using substantially all of the sound field signals corresponding to the current talkspurt, up to and including at least a first portion of the first time period.

7. (Previously Presented) A packet voice conferencing method comprising:

- receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;
- digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;
- estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
- transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
- wherein estimating the relative temporal delay comprises detecting the beginning time of a talkspurt in each of the sound field signals, and selecting the relative temporal delay for a talkspurt to correspond to the difference in beginning times detected for that talkspurt.

8. (Previously Presented) A packet voice conferencing method comprising:

- receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;

digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;

estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;

transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and

wherein the stereo decoding parameter expresses the estimated relative temporal delay between the first and second sound field signals as an integer number of digital sampling intervals.

9. (Original) The method of claim 1, wherein the stereo decoding parameter expresses an estimated angle of arrival based on the estimated relative temporal delay and the relative positioning of the first and second spatially-separated points.

10. (Previously Presented) A packet voice conferencing method comprising:

receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;

digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;

estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;

transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and

wherein the stereo decoding parameter corresponding to the digitally-encoded signal block representing the first time period is transmitted in the same packet as the digitally-encoded signal block.

11. (Previously Presented) A packet voice conferencing method comprising:

receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;

digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;

estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
wherein the stereo decoding parameter corresponding to the digitally-encoded signal block representing the first time period is transmitted in a later packet than the digitally-encoded signal block.

12. (Previously Presented) A packet voice conferencing method comprising:
receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;
digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;
estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
wherein the stereo decoding parameter corresponding to the digitally-encoded signal block representing the first time period is transmitted in a packet separate from any digitally-encoded signal block.

13. (Previously Presented) A packet voice conferencing method comprising:
receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;
digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;
estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
wherein the stereo decoding parameter is transmitted once per talkspurt.

14. (Previously Presented) A packet voice conferencing method comprising:

- receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;
- digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;
- estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
- transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
- estimating the signal energy present in each sound field signal during the approximate timeframe of the first time period, and transmitting to the remote conferencing endpoint, in packet format, an explicit stereo balance parameter related to the relative signal energy in each sound field signal.

15. (Previously Presented) A packet voice conferencing method comprising:

- receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;
- digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;
- estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;
- transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and
- estimating the signal energy present in a frequency subband of each sound field signal during the approximate timeframe of the first time period, and transmitting to the remote conferencing endpoint, in packet format, an explicit stereo balance parameter related to the relative signal energy in that subband for each sound field signal.

16. (Previously Presented) A packet voice conferencing method comprising:

- receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;

digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;

estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;

transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay; and

establishing a packet-based control protocol with the remote conferencing point, and using the control protocol to inform the remote conferencing point that an encoder performing the method of claim 1 is available for stereo packet voice conferencing.

17. - 46. (Canceled)

47. (Previously presented) A packet voice conferencing system comprising:

a packet parser to receive voice packets received from a remote conferencing point, each voice packet containing at least one of an encoded signal block and a stereo decoding parameter, the stereo decoding parameter comprising at least one of an explicit delay parameter, an explicit balance parameter, and an explicit arrival angle parameter;

a decoder to receive encoded signal blocks from the packet parser and decode those signal blocks to produce a voice sample stream; and

a playout splitter coupled to the voice sample stream, the splitter using the stereo decoding parameter to create multiple output signal channels based on the voice sample stream.

48. (Original) The packet voice conferencing system of claim 47, further comprising a jitter buffer inserted in the voice sample stream between the decoder and the playout splitter.

49. (Previously presented) The packet voice conferencing system of claim 47, wherein the stereo decoding parameter comprises an explicit delay parameter, the splitter delaying playout of the voice sample stream on at least one output signal channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the explicit delay parameter.

50. (Previously presented) The packet voice conferencing system of claim 47, wherein the stereo decoding parameter comprises an explicit balance parameter, the splitter modifying the

playout amplitude of the voice sample stream on at least one output signal channel, relative to the playout amplitude of the voice sample stream on another output signal channel, based on the value of the explicit balance parameter.

51. (Original) The packet voice conferencing system of claim 50, wherein the playout amplitude modification is audio-frequency dependent.

52. (Original) The packet voice conferencing system of claim 47, further comprising a mixer to mix the output signal channels with other signal channels derived from voice packets received from another remote conferencing point.

53. (Original) The packet voice conferencing system of claim 52, further comprising a packet formatter to place the mixer output in packet format for transmission to a remote conferencing endpoint.

54. (Previously presented) A packet voice conferencing system comprising:

means for decoding encoded signal blocks to produce a voice sample stream, each encoded signal block received in packet format from a remote conferencing point; and

means for splitting, based on the value of a stereo decoding parameter received in packet format from a remote conferencing point, the voice sample stream into multiple output signal channels to produce a stereophonic effect, the stereo decoding parameter comprising at least one of an explicit delay parameter, an explicit balance parameter, and an explicit arrival angle parameter.

55. (Previously presented) The packet voice conferencing system of claim 54, wherein the stereo decoding parameter comprises an explicit delay parameter, the means for splitting the voice sample stream comprising means for delaying playout of the voice sample stream on at least one output signal channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the explicit delay parameter.

56. (Previously presented) The packet voice conferencing system of claim 54, wherein the stereo decoding parameter comprises an explicit balance parameter, the means for splitting the voice sample stream comprising means for modifying the playout amplitude of the voice sample stream on at least one output signal channel, relative to the playout amplitude of the

voice sample stream on another output signal channel, based on the value of the explicit balance parameter.

57. (Previously presented) The packet voice conferencing system of claim 54, wherein the stereo decoding parameter comprises an explicit arrival angle parameter, the means for splitting the voice sample stream comprising means for calculating a delay parameter for at least one output signal channel to create the perception that the audio signal represented in the voice sample stream is arriving at an angle corresponding to the explicit arrival angle parameter.

58. (Previously presented) A packet voice conferencing method comprising:
receiving, from a remote conferencing point, a voice packet stream, at least some voice packets in the stream carrying a payload comprising an encoded signal block, at least some voice packets in the stream carrying a payload comprising a stereo decoding parameter, the stereo decoding parameter comprising at least one of an explicit delay parameter, an explicit balance parameter, and an explicit arrival angle parameter;
decoding the encoded signal blocks to produce a voice sample stream;
splitting the voice sample stream into multiple output signal channels; and
manipulating the signal carried on at least one of the output signal channels based on the value of the stereo decoding parameter to create a stereophonic effect on the output signal channels.

59. (Previously presented) The method of claim 58, wherein the stereo decoding parameter comprises an explicit delay parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises delaying playout of the voice sample stream on at least one output signal channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the explicit delay parameter.

60. (Previously presented) The method of claim 58, wherein the stereo decoding parameter comprises an explicit balance parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises modifying the playout amplitude of the voice sample stream on at least one output signal channel, relative to the playout amplitude of the voice sample stream on another output signal channel, based on the value of the explicit balance parameter.

61. (Previously presented) The method of claim 58, wherein the stereo decoding parameter comprises an explicit arrival angle parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises calculating a delay parameter for at least one output signal channel to create the perception that the audio signal represented in the voice sample stream is arriving at an angle corresponding to the explicit arrival angle parameter.

62. (Previously presented) An apparatus comprising a computer-readable medium containing computer instructions that, when executed, cause a processor or multiple communicating processors to perform a method for packet voice conferencing, the method comprising:

- receiving, from a remote conferencing point, a voice packet stream, at least some voice packets in the stream carrying a payload comprising an encoded signal block, at least some voice packets in the stream carrying a payload comprising a stereo decoding parameter, the stereo decoding parameter comprising at least one of an explicit delay parameter, an explicit balance parameter, and an explicit arrival angle parameter;

- decoding the encoded signal blocks to produce a voice sample stream;

- splitting the voice sample stream into multiple output signal channels; and

- manipulating the signal carried on at least one of the output signal channels based on the value of the stereo decoding parameter to create a stereophonic effect on the output signal channels.

63. (Previously presented) The apparatus of claim 62, wherein the stereo decoding parameter comprises an explicit delay parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises delaying playout of the voice sample stream on at least one output signal channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the explicit delay parameter.

64. (Previously presented) The apparatus of claim 62, wherein the stereo decoding parameter comprises an explicit balance parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises modifying the playout amplitude of the voice sample stream on at least one output signal channel, relative to the

playout amplitude of the voice sample stream on another output signal channel, based on the value of the explicit balance parameter.

65. (Previously presented) The apparatus of claim 62, wherein the stereo decoding parameter comprises an explicit arrival angle parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises calculating a delay parameter for at least one output signal channel to create the perception that the audio signal represented in the voice sample stream is arriving at an angle corresponding to the explicit arrival angle parameter.